

Citation for published version:

Chauvin, B, Gardy, D & Mailler, C 2015, 'A sprouting tree model for random boolean functions', *Random Structures and Algorithms*, vol. 47, no. 4, pp. 635-662. <https://doi.org/10.1002/rsa.20567>

DOI:

[10.1002/rsa.20567](https://doi.org/10.1002/rsa.20567)

Publication date:

2015

Document Version

Early version, also known as pre-print

[Link to publication](#)

This is the pre-peer reviewed version of the following article: Chauvin, B, Gardy, D & Mailler, C 2015, 'A sprouting tree model for random boolean functions' *Random Structures and Algorithms*, vol 47, no. 4, pp. 635-662., , which has been published in final form at <http://dx.doi.org/10.1002/rsa.20567>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A sprouting tree model for random Boolean functions.*

Brigitte Chauvin[†], Danièle Gardy[‡] and Cécile Mailler[§]

June 9, 2014

Abstract

We define a new probability distribution for Boolean functions of k variables. Consider the random Binary Search Tree of size n , and label its internal nodes by connectives and its leaves by variables or their negations. This random process defines a random Boolean expression which represents a random Boolean function. Finally, let n tend to infinity: the asymptotic distribution on Boolean functions exists; we call it the *sprouting tree* distribution. We study this model and compare it with two previously-known distributions induced by two other random trees: the Catalan tree and the Galton-Watson tree.

Keywords

Boolean functions, Boolean expressions, Boolean formulas, binary trees, Binary Search Tree model of growth, Yule tree.

*This work was partially supported by the ANR project 09 BLAN 0011 Boole.

[†]Laboratoire de Mathématiques de Versailles. Université de Versailles St-Quentin en Yvelines, 45 avenue des Etats-Unis, 78035 Versailles, France. Email: brigitte.chauvin@math.uvsq.fr

[‡]Laboratoire PRISM. Université de Versailles St-Quentin en Yvelines, 45 avenue des Etats-Unis, 78035 Versailles, France. Email: danièle.gardy@prism.uvsq.fr.

[§]Laboratoire de Mathématiques de Versailles et Prob@LaB, University of Bath, Claverton Down, BA2 7AY Bath, UK. Email: c.mailler@bath.ac.uk.

1 Introduction

Consider the set of Boolean functions on k variables: which probability models on this set are reasonable and meaningful? The first, natural answer to such a question is to consider the uniform distribution on this finite set. This distribution was already studied some seventy years ago by Riordan and Shannon [RS42] who were interested in the *complexity* of an average Boolean function, the complexity of a Boolean function being the minimal number of logical connectives (And and Or connectives for example) needed to represent the Boolean function by a Boolean formula (see [Weg05] for an introduction to the complexity of Boolean functions). According to this uniform distribution, and asymptotically when the number of variables k tends to infinity, *almost all Boolean functions are of almost maximal complexity*; this property is called the “Shannon effect”.

Much later, Paris *et al.* [PVW94], Lefmann and Savický [LS97], Chauvin *et al.* [CFGG04] and others considered distributions induced by the tree representation of Boolean functions (a survey can be found in [Gar06]). Boolean formulas can be represented by labelled binary complete trees, where internal nodes (vertices with degree 2 or 3) are labelled by logical connectives and leaves (vertices with degree 1) by literals, i.e. variables or their negations. Each such tree computes (represents) a Boolean function. Therefore any probability distribution over the set of trees induces a specific probability distribution over the set of Boolean functions. However, as several trees can represent the same Boolean function, the properties of the probability distribution induced on the set of Boolean functions are not easily seen.

Let us expand this approach. Defining a distribution induced by a random tree representation amounts to choosing both a logical system (i.e. a set of allowed labels) and a distribution over the set of labelled trees. Common logical systems are, e.g., the And/Or model (we allow the And and Or connectives, and positive or negative literals), and the implication model (the only allowed connective is the implication, and only positive literals are allowed). The And/Or model is complete, i.e. every Boolean function can be represented in this model; it was studied for example in [PVW94, LS97, CFGG04]. The implication model is not complete; it was nevertheless studied because of its logical properties and relation to intuitionistic logic in [MTZ00, FGGZ07, KZ04, GK09] among others.

Consider first And/Or trees. Lefmann and Savický ([LS97]) and Chauvin *et al.* ([CFGG04]) have studied the distribution on Boolean functions stemming from the uniform distribution over the set of all trees of a given size (the size of a tree being in the present paper the number of its internal nodes). Pick a tree at random according to this distribution, then label each of its nodes independently and uniformly, i.e. with And connectives or Or connectives chosen with uniform probability at each internal node, and with literals uniformly chosen among $\{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$ at each leaf. First Lefmann and Savický, then Chauvin *et al.*, proved that the distribution induced on Boolean functions by such a labelling admits a limiting distribution when the size of the trees tends to infinity; this asymptotic distribution is often called the *Catalan tree* distribution. Chauvin *et al.* studied it through analytic combinatorics tools (see, e.g., [FS09] for an introduction to this method) and showed that the Catalan tree distribution *gives a larger probability to low complexity functions than to high complexity functions*. This result has been sharpened by Kozik [Koz08] who used a very general and powerful approach through pattern description: the Catalan tree distribution on Boolean functions of k variables deeply differs from the uniform law. Kozik also gives a very precise description of the average (according to the Catalan tree distribution) tree computing a given Boolean function f : this is a tree computing f and of minimal size, on which has been *grafted* one large tree of a specific shape. Note however that the results of Kozik are valid asymptotically when the number k of Boolean variables tends to infinity, and are not as general as the results of Lefmann and Savický and Chauvin *et al.*

Chauvin *et al.* also define another distribution on trees, which in turn leads to another distribution on Boolean functions, through a stochastic branching process. Consider the binary critical Galton-Watson process: it is known that this process ends almost surely. Label the resulting random tree uniformly at random, just as for the Catalan tree. The result is a random labelled tree; its probability distribution again induces a distribution over the set of Boolean functions, which goes by the name of *Galton-Watson tree* distribution. Just as the Catalan tree distribution did, it gives a comparatively larger probability to low complexity functions [CFGG04].

Concurrently with these results on And/Or trees, similar results were obtained for the implication model [FGGG08]. Then it was proved by Genitrini and Gittenberger ([GG10]) that neither of the Cata-

lan tree nor the Galton-Watson tree distributions exhibit a Shannon effect. A large part of the studies on the implication model has also investigated tautologies (trees computing the constant function *True*), thus providing a tool to compare intuitionistic and classical logic [GK09]: asymptotically when the number of variables k tends to infinity, almost every tautology has a very simple shape, with respect to both the Catalan tree distribution and the Galton-Watson tree distribution.

In the present paper, we introduce a new probability distribution induced by the tree representation of Boolean functions: the *sprouting tree* distribution. We build a random tree of size n as the result of a stochastic process inspired from the growing process of a random Binary Search Tree: *the tree grows uniformly from its leaves*. Label it uniformly (in the same way as for the Catalan and Galton-Watson models): we define a distribution over the set of labelled trees of size n . We show in this paper that in both the And/Or and the implication models, the sequence of distributions induced on Boolean functions by such random trees admits a limit when the size of the trees tends to infinity, which we call the *sprouting tree* distribution.

Our new model is equivalent to building a random Boolean expression in a very natural way, as follows: start with an expression which is simply a literal, then extend it step by step, by picking up uniformly at random a literal from the expression, then replacing it by a random sub-expression of the shape $\ell_1 \diamond \ell_2$, where ℓ_1 and ℓ_2 are two literals and \diamond is a connective.

The underlying tree (the Binary Search Tree) is more *balanced* than either the Catalan or the Galton-Watson trees: its expected height is of order $\ln n$, while the expected height of both the Catalan and Galton-Watson trees is of order \sqrt{n} . This gives a hint that we may expect the behaviour of the induced distribution on Boolean functions to differ from the previously-known ones.

The paper is organized as follows: we sum up in Section 2 the results gathered in the literature about the uniform law (studied by Riordan and Shannon), the Catalan tree distribution and the Galton-Watson tree distribution. Then we state and prove in Section 3 the main result of this paper: the sprouting tree distribution gives probability zero to non-constant functions. Next we focus on the And/Or model and study the effect of relaxing the uniform labelling assumption in Section 4. Section 5 deals with the implication model: we prove there that, according to the sprouting tree distribution, almost every tautology is “non simple” asymptotically when k tends to infinity. A table summarizing the main notations of the paper and comparing the sprouting tree model with Catalan tree and Galton-Watson tree models can be found in Section 6 (see Table 2). We also refer the reader to Table 1 for a reminder of some notations used in the proofs.

2 Tree-induced distributions on Boolean functions

Let \mathcal{F}_k be the set of Boolean functions over k variables, i.e. of functions f from the set $\{0, 1\}^k$ to the set $\{0, 1\}$. In all the paper, (x_1, \dots, x_k) is an element of $\{0, 1\}^k$ and $f(x_1, \dots, x_k)$ is the value of f at (x_1, \dots, x_k) . Moreover, the values 0 and 1 will often be replaced by their truth values, respectively *False* and *True*.

2.1 Context

The set \mathcal{F}_k is finite, of cardinality 2^{2^k} . Therefore the most basic idea, when attempting to define a probability distribution over Boolean functions, is simply to give equal probability to all functions, i.e. to consider the uniform distribution over \mathcal{F}_k . This distribution has been studied, among others, by Riordan and Shannon [RS42] who have shown the so-called *Shannon effect*: almost all Boolean functions have almost maximal complexity – the *complexity* of a Boolean function f in a logical system being the minimal number of connectives needed to represent it by a Boolean expression in this model (see [Weg05]).

Example: In the And/Or model, the Boolean function XOR can be defined as $x_1 \text{ XOR } x_2 = (\bar{x}_1 \wedge x_2) \vee (x_1 \wedge \bar{x}_2)$, and has complexity 3 (cf. Figure 1(A)).

The following theorem sums up classical results relative to the complexity of a Boolean function.

Theorem 1. (i) *The maximal complexity of a Boolean function on k variables is of order $\frac{2^k}{\log_2 k} (1 + o(1))$ when k tends to infinity.*

(ii) When the set \mathcal{F}_k is equipped with the uniform distribution, almost every Boolean function of k variables has complexity larger than $\frac{(1-\varepsilon)2^k}{\log_2 k}$.

Remark: Statement (i) in Theorem 1 is due to Lupanov [Lup60], we refer to Jukna [Juk12, page 32] for more details on this bound. Statement (ii) is due to Riordan and Shannon [RS42], we refer to [FS09, page 78] for a simple proof.

Another way to define a probability distribution over the set \mathcal{F}_k of Boolean functions is to consider tree representations of Boolean functions. A Boolean function can be represented by many Boolean expressions, which we identify with labelled binary plane rooted trees where the leaves are labelled either by a variable (x_i) or by the negation of a variable (\bar{x}_i) and the internal nodes by connectives (e.g. $\wedge, \vee, \rightarrow$).

Choosing the rules that define the legal labellings of trees amounts to choosing a logical system. For example, if we allow the connectives \wedge (And) and \vee (Or), and if a leaf can be labelled by a variable or by the negation of a variable, then we consider the so-called And/Or model. This model is complete, which means that every Boolean function can be represented by such a tree; it has therefore been studied in different papers, e.g. [LS97, CFGG04]. Figure 1(A) represents an And/Or tree. We shall consider in this paper both the And/Or model and the *implication model*, which we define as follows. The implication is the single connective (\rightarrow ¹) that can be used as a label for internal nodes; the labels for the leaves are restricted to the Boolean variables (not their negations). This logical system is not complete, e.g. the function *False* cannot be represented by such a tree, but it is simple (only one connective) and has interesting logical properties (see [MTZ00, FGGZ07, GK09] for example, where this logical system has been used to prove that, roughly speaking, *classical and intuitionistic logic are asymptotically the same*). An example of an implication tree is given in Figure 1(B).

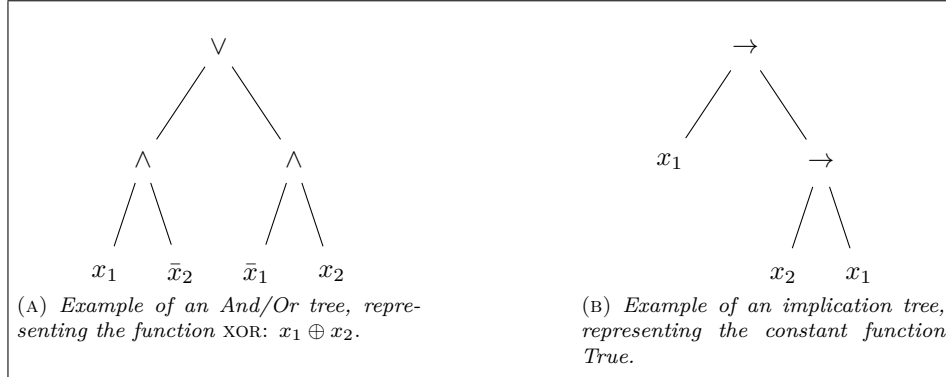


FIGURE 1 – Examples of tree representations

Definition 1. The size of a tree is the number of its internal nodes.²

Definition 2. The complexity $L(f)$ of a Boolean function f is

- the size of a minimal tree representing f , if f is a non-constant function;
- 0 if f is one of the two constant functions.

Definition 3. Let \mathcal{E}_k be the set of labelled binary trees on k Boolean variables (either And/Or trees or implication trees in this paper). Let Φ be the function defined as follows:

$$\begin{aligned} \Phi : \mathcal{E}_k &\rightarrow \mathcal{F}_k \\ t &\mapsto \Phi(t) = f \text{ if and only if } t \text{ computes } f. \end{aligned}$$

¹Recall that the implication is defined by: $x \rightarrow y \equiv y \vee \bar{x}$.

²In the literature, the size also appears as the number of leaves of the trees; in our model, counting internal node is more appropriate for the calculations. Of course, since a tree with n internal nodes has $n + 1$ leaves, this choice makes no difference whatsoever.

If ν is a probability measure on \mathcal{E}_k , the image measure μ of ν by Φ is defined by $\mu(f) = \nu(\Phi^{-1}(f)) = \nu(\{t \text{ such that } t \text{ represents } f\})$.

In the following, we describe three distributions over \mathcal{F}_k induced from the tree representation. We already alluded to results obtained on the Catalan tree and the Galton-Watson tree distributions; see Lefmann and Savický [LS97], Chauvin *et al.* [CFGG04], Kozik [Koz08], Genitrini and Gittenberger [GG10], and Fournier *et al.* [FGGG12]. We first sum up their results, before considering a new model of random trees which induces a new distribution over \mathcal{F}_k .

2.2 The Catalan tree

Pick a tree uniformly at random from the set of unlabelled trees of size n , then label it uniformly at random (i.e. each internal node chooses its label uniformly among the allowed connectives and each leaf chooses its label uniformly among the set of allowed literals, all the choices being independent from each other); denote by $\mathbb{U}_{n,k}$ the distribution of the random labelled tree we obtain. This distribution induces (through Φ) a distribution $\mu_{n,k}$ over the set \mathcal{F}_k . The following theorems define and describe the Catalan tree distribution (see Table 2 for a summary of these results).

Theorem 2 (Lefmann and Savický [LS97], Chauvin *et al.* [CFGG04]). *In the And/Or model, the sequence $(\mu_{n,k})_{n \geq 0}$ of probability distributions over \mathcal{F}_k converges to a probability distribution μ_k when n tends to infinity. The distribution μ_k is called the Catalan tree distribution.*

Theorem 3 (Kozik [Koz08] in the And/Or model, Fournier *et al.* [FGGG12] in the implication model). *Let k_0 be an integer. Given a fixed Boolean function $f \in \mathcal{F}_{k_0}$, we have that*

$$\mu_k(f) = \Theta\left(\frac{1}{k^{L(f)+2}}\right), \text{ when } k \text{ tends to infinity.}$$

Remark:

- The Boolean function f in Theorem 3 depends on a fixed number k_0 of variables, but the total number k of Boolean variables tends to infinity.
- In Theorem 3, we let first the size n of the trees tend to infinity, *then* the number k of variables used for the labelling: the order is important.

The Catalan tree distribution behaves differently from the uniform distribution studied by Shannon:

Theorem 4 (Genitrini and Gittenberger [GG10]). *The Catalan tree distribution presents no Shannon effect in the implication model (the set of functions with complexity of order k^2 has a positive asymptotic distribution probability when k tends to infinity).*

2.3 The Galton-Watson tree

A second distribution induced by trees over the set of Boolean functions on k variables is inspired by the critical binary Galton-Watson process, and was introduced and studied by Chauvin *et al.* [CFGG04]. The critical binary Galton-Watson process is a stochastic process of non-labelled binary trees defined as follows: at time 0, the Galton-Watson tree is a single vertex; at time $n > 0$, we consider each leaf of the n -th generation independently from the others and each leaf dies with no child with probability $1/2$ or gives birth to two children with probability $1/2$. It is known that this process ends almost surely; therefore this stochastic process defines a finite random binary tree.

Labelling this random tree *uniformly* at random defines a probability distribution Π_k over \mathcal{E}_k , which induces by Φ a probability distribution π_k over \mathcal{F}_k (see Definition 3).

Results in the vein of Theorems 3 and 4 also hold for the Galton-Watson tree distribution, as proved in [FGGG12, GG10] (we refer to Table 2 for a summary of these results).

2.4 A new model: the sprouting tree

The model we introduce in this paper is inspired from the random Binary Search Trees (BST) growing process (see [CLR89, page 254] for a description of the random Binary Search Tree). As for the Catalan and Galton-Watson tree distributions, we first define a random unlabelled tree, called the sprouting tree, by some inductive stochastic process, before labelling it uniformly at random to define a probability distribution over \mathcal{F}_k . The sprouting tree grows uniformly from its leaves:

Definition 4 ([CKMR05]). *The sprouting tree $(\mathcal{T}_i)_{i \in \mathbb{N}}$ is a sequence of random trees defined by*

- \mathcal{T}_0 is a single vertex.
- Given \mathcal{T}_i , we choose uniformly at random a leaf of the tree and make it grow by giving it two children. The new tree is \mathcal{T}_{i+1} .

The random tree \mathcal{T}_n is called the sprouting tree of size n .

After labelling \mathcal{T}_n uniformly at random, we obtain a labelled random tree $\mathcal{T}_{n,k}$ (where k is the number of variables used in the labelling), still called the sprouting tree for simplicity's sake, which defines a probability distribution $\mathbb{P}_{n,k}$ over \mathcal{E}_k .

Definition 5. *The distribution induced on \mathcal{F}_k by $\mathbb{P}_{n,k}$ through Φ is denoted by $p_{n,k}$: for all Boolean functions $f \in \mathcal{F}_k$,*

$$p_{n,k}(f) = \mathbb{P}(\mathcal{T}_{n,k} \text{ computes } f).$$

The aim of this paper is to answer some natural questions: does the sequence $(p_{n,k})_{n \geq 0}$ of probability distributions have a limit when n tends to infinity? If it does, what can be said about the asymptotic distribution? Can we prove or disprove the Shannon effect?

3 The sprouting tree distribution

3.1 Existence of the sprouting tree distribution

The following theorems show that the asymptotic distribution p_k of the $(p_{n,k})_{n \geq 0}$ exists; we call it the *sprouting tree distribution*. Moreover, this new distribution over \mathcal{F}_k is surprisingly simple; it gives probability zero to all non-constant functions.

Theorem 5 (Sprouting tree - And/Or model). *For the And/Or labelling model and when n tends to infinity, we have that $p_{n,k} \rightarrow p_k = \frac{1}{2}\delta_{True} + \frac{1}{2}\delta_{False}$ where, for all $f \in \mathcal{F}_k$, δ_f is defined as the probability distribution on \mathcal{F}_k such that $\delta_f(f) = 1$. Moreover, $\|p_{n,k} - p_k\|_\infty = \mathcal{O}\left(\frac{1}{\ln n}\right)$ when n tends to infinity³.*

We shall see in Section 4 that Theorem 5 can be extended to more general labelling models.

Theorem 6 (Sprouting tree - Implication model). *In the case of the implication labelling model and when $n \rightarrow +\infty$, we have that $p_{n,k} \rightarrow p_k = \delta_{True}$. Moreover, $\|p_{n,k} - p_k\|_\infty = \mathcal{O}\left(\frac{1}{\ln n}\right)$ when n tends to infinity.*

The two theorems are slightly different, since the sprouting tree distribution gives probability 1 to *True* in the implication model, while it gives probability $1/2$ to *False* and *True* in the And/Or model. This difference is easily explained: the function *False* cannot be represented in the implication logical system, while the And/Or system is complete.

The proofs of these two theorems are very similar; therefore we develop only the And/Or case, which is the more complicated of the two. We present two different proofs of the theorem: an analytic combinatorics proof, which is natural since it was the approach used in the previous models, and a probabilistic proof based on the study of Yule trees, which is specific to this particular model.

The following table summarizes the notations used in both proofs:

³In the present paper, the norm $\|\cdot\|_\infty$ is defined on the set of signed measures on \mathcal{F}_k as follows: for all signed measure p , $\|p\|_\infty = \sup_{f \in \mathcal{F}_k} |p(f)|$.

	Non labelled tree	Labelled tree	Law of the labelled tree	Induced law on \mathcal{F}_k	Asymptotic distribution
Discrete time	\mathcal{T}_n	$\mathcal{T}_{n,k}$	$\mathbb{P}_{n,k}$	$p_{n,k}$	p_k
Continuous time	\mathcal{Y}_t	$\mathcal{Y}_{t,k}$	$\mathbb{P}_{t,k}$	$\mathbf{p}_{t,k}$	p_k

TABLE 1 – Discrete and continuous time models: notations

3.2 An analytic combinatorics proof

The idea here is to use generating functions and analytic combinatorics methods, as presented, e.g., by Flajolet and Sedgewick [FS09]. Given a sequence $(s_n)_{n \geq 0}$ (for example $(p_{n,k}(f))_{n \geq 0}$), consider the power series, also called the generating function of $(s_n)_{n \geq 0}$, $S(z) = \sum_{n \geq 0} s_n z^n$. The behaviour of the power series, seen as an analytic function near its dominant singularity, can give some useful information about the asymptotic behaviour of the initial sequence.

Definition 6. Given a Boolean function f in \mathcal{F}_k , define its generating function $\phi(f)$ by

$$\phi_f(z) = \sum_{n=0}^{+\infty} p_{n,k}(f) z^n,$$

where $p_{n,k}(f)$ is the probability that the random sprouting tree $\mathcal{T}_{n,k}$ of size n computes f . We denote $\phi_T(z) := \Phi_{True}(z)$.

We shall need the following, standard result about the sprouting tree:

Lemma 1. For all $n \geq 1$, the subtrees of $\mathcal{T}_{n,k}$, the sprouting tree of size n , are themselves sprouting trees, and the probability that the left subtree has a given size $k \in \{0, \dots, n-1\}$ is $1/n$.

Proof. This result is standard but we include its proof for the sake of completeness. We can prove Lemma 1 by induction; if $n = 1$, then the left subtree has size 0 with probability 1. Fix $n \geq 1$ and assume that the size of the left subtree⁴ of $\mathcal{T}_{n,k}$, denoted by \mathcal{L}_n , follows the uniform distribution over $\{0, \dots, n-1\}$. If we obtain a left subtree of size i at step $n+1$, either the left subtree already had size i at step n and the sprouting leaf was chosen in the right subtree, or the left subtree had size $i-1$ at step n and the sprouting leaf was chosen in the left subtree. If we condition by the event “at the n th step, the left subtree had size $i-1$ ” or “at the n th step, the left subtree had size i ”, we obtain, for all $i \in \{1, \dots, n-1\}$,

$$\begin{aligned} \mathbb{P}(|\mathcal{L}_{n+1}| = i) &= \left(1 - \frac{i+1}{n+1}\right) \mathbb{P}(|\mathcal{L}_n| = i) + \frac{i}{n+1} \mathbb{P}(|\mathcal{L}_n| = i-1) \\ &= \frac{n-i}{n+1} \cdot \frac{1}{n} + \frac{i}{n+1} \cdot \frac{1}{n} = \frac{1}{n+1}. \end{aligned}$$

Moreover,

$$\mathbb{P}(|\mathcal{L}_{n+1}| = 0) = \frac{n}{n+1} \mathbb{P}(|\mathcal{L}_n| = 0) = \frac{1}{n+1},$$

and,

$$\mathbb{P}(|\mathcal{L}_{n+1}| = n) = \frac{n}{n+1} \mathbb{P}(|\mathcal{L}_n| = n-1) = \frac{1}{n+1}.$$

Thus, the size of the left subtree \mathcal{L}_{n+1} of $\mathcal{T}_{n+1,k}$ follows the uniform law over $\{0, \dots, n\}$. □

We are now able to state the following result:

⁴The sizes of the left and right subtrees of a tree of size n are linked: their sum is equal to $n-1$.

Lemma 2. For all $f \in \mathcal{F}_k$, its generating function (see Definition 6) $\phi_f(z)$ satisfies

$$\begin{aligned} 2\phi'_f(z) &= \sum_{g \wedge h = f} \phi_g(z)\phi_h(z) + \sum_{g \vee h = f} \phi_g(z)\phi_h(z) \\ \phi_f(0) &= \frac{1}{2k} \mathbb{1}_{\{f \text{ lit}\}}, \end{aligned}$$

where $\mathbb{1}_{\{f \text{ lit}\}} = 1$ when there exists $i \in \{1, \dots, k\}$ such that $f : (x_1, \dots, x_k) \mapsto x_i$ or $f : (x_1, \dots, x_k) \mapsto \bar{x}_i$.

Proof. The sprouting tree $\mathcal{T}_{n,k}$ of size n computes f if and only if

- $n = 0$, $f = \alpha$ is a literal, and the root of $\mathcal{T}_{0,k}$ (which is also its unique leaf) is labelled by α ; or
- $n \geq 1$, the left subtree of $\mathcal{T}_{n,k}$ computes g , the right subtree of $\mathcal{T}_{n,k}$ computes h , the root of $\mathcal{T}_{n,k}$ is labelled by $\diamond \in \{\wedge, \vee\}$ and $f = g \diamond h$.

Therefore, in view of Lemma 1, by conditioning on the size of the left subtree, we obtain the following formula for all $n \geq 0$,

$$p_{n+1,k}(f) = \frac{1}{2} \sum_{g \wedge h = f} \sum_{i=0}^n \frac{1}{n+1} p_{i,k}(g) p_{n-i,k}(h) + \frac{1}{2} \sum_{g \vee h = f} \sum_{i=0}^n \frac{1}{n+1} p_{i,k}(g) p_{n-i,k}(h). \quad (1)$$

Multiplying (1) by $(n+1)z^n$ and summing for $n \geq 0$ gives Lemma 2. \square

Lemma 2 gives a system of first-order ordinary differential equations satisfied by the 2^{2^k} generating functions for the 2^{2^k} Boolean functions of \mathcal{F}_k . Studies of the Catalan tree and of the Galton-Watson tree distributions by this method lead to very similar systems, except that they are both algebraic systems [CFGG04]. In those cases, it was straightforward to conclude through the Drmota-Lalley-Woods theorem, which applies to algebraic systems [Drm97, Lal93, Woo97] (see [FS09, page 489]). In our case, the appearance of a derivative forbids applying this theorem. Fortunately, the asymptotic behaviour of any solution of the system obtained from Lemma 2 can still be derived as follows.

We first state that obvious symmetries simplify the system. The uniform labelling ensures us that for all Boolean function $f \in \mathcal{F}_k$, for all $n \geq 0$, $p_{n,k}(f) = p_{n,k}(\bar{f})$, which implies $\phi_f(z) = \phi_{\bar{f}}(z)$ for all $f \in \mathcal{F}_k$.

Proposition 1. There exists a constant $\kappa > 0$ such that, for all $z \in [0, 1)$,

$$\frac{1}{2(1-z)} - \frac{1/\kappa}{2(1-z) \left(1/\kappa + \ln \left(\frac{1}{1-z} \right) \right)} \leq \phi_T(z) \leq \frac{1}{2(1-z)}.$$

Proof. Let us denote by

$$\phi_S(z) = \sum_{f \notin \{\text{True}, \text{False}\}} \phi_f(z).$$

In all this proof, we take $z \in [0, 1[$. Remark that

$$\phi_S + 2\phi_T = \frac{1}{1-z},$$

and, in view of Lemma 2,

$$\begin{aligned} 2\phi'_S(z) &= \sum_{f \notin \{\text{True}, \text{False}\}} \sum_{g \wedge h = f} \phi_g(z)\phi_h(z) + \sum_{f \notin \{\text{True}, \text{False}\}} \sum_{g \vee h = f} \phi_g(z)\phi_h(z) \\ &\leq 2 \sum_{g, h \notin \{\text{True}, \text{False}\}} \phi_g(z)\phi_h(z) \\ &\quad - 2 \sum_{g \notin \{\text{True}, \text{False}\}} \phi_g(z)\phi_{\bar{g}}(z) \\ &\quad + 4\phi_T(z) \sum_{g \notin \{\text{True}, \text{False}\}} \phi_g(z), \end{aligned}$$

because a non-constant function is

- either the conjunction or disjunction of two non-constant functions (first term of the above sum),
- except that every $g \vee \bar{g}$ or $g \wedge \bar{g}$ is a constant function (second term of the above sum),
- or the conjunction of a non-constant function with *True*, or the disjunction of a non-constant function with *False* (third term of the above sum).

Note that the previous formula is an inequality because the right term includes conjunctions like $x \wedge (\bar{x} \wedge y) \equiv \text{False}$ which are constant functions. We thus have:

$$\begin{aligned} 2\phi'_S(z) &\leq 2 \sum_{g,h \notin \{True, False\}} \phi_g(z)\phi_h(z) - 2 \sum_{f \notin \{True, False\}} \phi_f(z)\phi_{\bar{f}}(z) + 4\phi_S(z)\phi_T(z) \\ &\leq 2\phi_S(z)^2 + 4\phi_S(z)\phi_T(z) - 2\kappa\phi_S(z)^2, \end{aligned}$$

where $\kappa > 0$ is a constant such that

$$\sum_{f \notin \{True, False\}} \phi_f(z)^2 \geq 2\kappa\phi_S(z)^2.$$

The existence of the constant κ results from the convexity of the function $x \mapsto x^2$. Hence

$$\phi'_S(z) + \kappa\phi_S(z)^2 \leq 2\phi_S(z)\phi_T(z) + \phi_S(z)^2 = \frac{\phi_S(z)}{1-z}.$$

Now let $Y(z)$ be the solution of $Y'(z) + \kappa Y(z)^2 = \frac{Y(z)}{1-z}$ such that $Y(0) = 1$. This equation is a Bernoulli differential equation that we can solve through a standard change of variable. Let $W(z) = \frac{1}{Y(z)}$ and $\psi(z) = \frac{1}{\phi_S(z)}$; these functions satisfy the equation

$$W'(z) + \frac{W(z)}{1-z} = \kappa$$

and

$$\psi'(z) + \frac{\psi(z)}{1-z} \geq \kappa,$$

with initial conditions $W(0) = \psi(0) = 1$. Thus, via Grönwall's lemma [Grö19], for all $z \in [0, 1)$,

$$\psi(z) \geq W(z).$$

Moreover,

$$W(z) = \kappa(1-z) \left(\frac{1}{\kappa} + \ln \left(\frac{1}{1-z} \right) \right),$$

which implies, for all $z \in [0, 1)$,

$$\phi_S(z) \leq \frac{1/\kappa}{(1-z) \left(\frac{1}{\kappa} + \ln \left(\frac{1}{1-z} \right) \right)}.$$

□

This asymptotic behaviour only holds on the real line. Consequently we cannot use the classical transfer lemma of Flajolet and Odlyzko [FO90] detailed in [FS09, page 389]. However, a standard Tauberian theorem (see for example [Har49, page 155]) allows us to obtain the asymptotic behaviour of the partial sums of the coefficients $p_{n,k}(f)$ of the generating functions $\phi_f(z)$: when n tends to infinity,

$$\sum_{i=1}^n \left(\frac{1}{2} - p_{i,k}(True) \right) = \sum_{i=1}^n \left(\frac{1}{2} - p_{i,k}(False) \right) = \mathcal{O} \left(\frac{n}{\ln n} \right). \quad (2)$$

We are now in a position to obtain the behaviour of the coefficients of $\phi_T(z)$ when n tends to infinity.

In view of Equation (1),

$$\begin{aligned}
p_{n+1,k}(True) &= \frac{1}{2} \sum_{g \wedge h = True} \sum_{i=0}^n \frac{1}{n+1} p_{i,k}(g) p_{n-i,k}(h) + \frac{1}{2} \sum_{g \vee h = True} \sum_{i=0}^n \frac{1}{n+1} p_{i,k}(g) p_{n-i,k}(h) \\
&\geq \frac{1}{2(n+1)} \sum_{i=1}^n p_{i,k}(True) p_{n-i,k}(True) \\
&\quad + \frac{1}{n+1} \sum_{i=1}^n p_{i,k}(True) (1 - p_{n-i,k}(True)) + \frac{1}{2(n+1)} \sum_{i=1}^n p_{i,k}(True) p_{n-i,k}(True)
\end{aligned}$$

where the first term of the last sum stands for the probability that $\mathcal{T}_{n,k}$ represents *True* when its root is labelled by \wedge , and the sum of the two other terms stands for the same probability when the root is labelled by \vee . Therefore, Equation (2) implies

$$\frac{1}{2} - p_{n+1,k}(True) \leq \frac{1/2}{n+1} + \frac{1}{n+1} \sum_{i=1}^n \left(\frac{1}{2} - p_{i,k}(True) \right) = \mathcal{O} \left(\frac{1}{\ln n} \right), \text{ when } n \rightarrow +\infty,$$

which proves Theorem 5.

3.3 A probabilistic proof

We use here an idea due to Pittel for Binary Search Trees [Pit84]: we embed the discrete process of the sprouting tree into a continuous time process by way of “exponential clocks”, i.e. exponentially distributed random variables.

Instead of growing step by step at times $(1, 2, \dots, n, \dots)$, the tree now grows at random continuous times: each leaf sprouts independently from the others and after an exponentially distributed time. We thus define a continuous process of trees, denoted by $(\mathcal{Y}_t)_{t \geq 0}$ and called the *Yule tree* (see Definition 7). The link with the (discrete) sprouting tree is as follows: if we consider the sequence of the distinct trees which are the values of the continuous process, then this sequence is a sprouting tree. This property comes from the fact that the times of growth are independent and exponentially distributed. Thus, studying the continuous time process will give information about the (discrete) sprouting tree.

Moreover, the Yule tree has a useful property which justifies the continuous time embedding, as it does not hold in discrete time: the right and left subtrees at each node of the tree are independent;. This property is the key to our proof of Theorems 5 and 6.

Definition 7 (Pittel [Pit84]). *A Yule tree is a continuous time process of binary trees $(\mathcal{Y}_t)_{t \geq 0}$ growing according to the following rules:*

- \mathcal{Y}_0 is a single root;
- each leaf of \mathcal{Y}_t gives birth to two children at the end of a random time following an exponential law of parameter 1, independently from the other leaves.

Definition 8. *A labelled Yule tree is a continuous time process $(\mathcal{Y}_{t,k})_{t \geq 0}$ of labelled binary trees, which evolves according to the following rules:*

- the underlying binary tree is a Yule tree;
- each new leaf is labelled by a literal chosen uniformly at random among $\{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$;
- each new internal node is labelled by \wedge or \vee uniformly at random;
- each labelling is independent from the others.

Let us denote by $\mathbf{P}_{t,k}$ the law of $\mathcal{Y}_{t,k}$ and by $\mathbf{p}_{t,k}$ its image by Φ .⁵

⁵Let us recall that Φ is the surjective mapping from \mathcal{E}_k to \mathcal{F}_k such that $\Phi(\gamma) = f$ if and only if γ represents (or computes) f .

Definition 9. For all Boolean functions f ,

$$\mathbf{p}_{t,k}(f) = \mathbb{P}(\mathcal{Y}_{t,k} \text{ computes } f).$$

Fact: For all $t \geq 0$, let us denote by $N(t)$ the number of internal nodes of $\mathcal{Y}_{t,k}$. Then, $\mathcal{Y}_{t,k}$ has the same distribution as $\mathcal{T}_{N(t),k}$; it is a sprouting tree of size $N(t)$.

The main idea to prove Theorem 5 is inspired by an article about fully balanced binary trees [FGG09]: we consider the probability that two different assignments have distinct images under a random Boolean function, and prove that this probability tends to 0 as t tends to infinity. Therefore, the asymptotic distribution gives weight zero to non-constant functions.

Let $a = (a_1, \dots, a_k)$ and $b = (b_1, \dots, b_k)$ be two distinct elements of $\{0, 1\}^k$, i.e. two different assignments of the k variables. Let α and β be two elements of $\{0, 1\}$ and $f \in \mathcal{F}_k$ a Boolean function. For all $t \geq 0$, we denote $\mathbf{p}_{t,k}^{\alpha\beta}(a, b) = \mathbf{p}_{t,k}(f(a) = \alpha \text{ and } f(b) = \beta)$.

Fact: The symmetries between \wedge and \vee and between the variables and their negations imply that $\mathbf{p}_{t,k}^{01}(a, b) = \mathbf{p}_{t,k}^{10}(a, b)$ and $\mathbf{p}_{t,k}^{00}(a, b) = \mathbf{p}_{t,k}^{11}(a, b)$. Indeed, the probabilities that a random tree computes either f or its negation \bar{f} are the same, since \wedge and \vee occur with the same probability at each internal node, and a variable or its negation occur with the same probability at each leaf. Moreover, $\mathbf{p}_{t,k}^{10}(a, b) + \mathbf{p}_{t,k}^{00}(a, b) = 1/2$ for all a and b in $\{0, 1\}^k$.

Conditioning on the time when the root gives birth to two children (the time follows an exponential law of parameter 1) and writing $\mathbf{p}_t^{\alpha\beta}$ instead of $\mathbf{p}_{t,k}^{\alpha\beta}(a, b)$ to simplify notations when there is no ambiguity, we obtain that

$$\mathbf{p}_t^{10} = \sum_{i=1}^k \frac{e^{-t}}{2k} \mathbb{1}_{\{a_i \neq b_i\}} + \frac{1}{2} \int_0^t (\mathbf{p}_{t-s}^{11} \mathbf{p}_{t-s}^{10} + \mathbf{p}_{t-s}^{10} (\mathbf{p}_{t-s}^{11} + \mathbf{p}_{t-s}^{10}) + \mathbf{p}_{t-s}^{10} (\mathbf{p}_{t-s}^{00} + \mathbf{p}_{t-s}^{10}) + \mathbf{p}_{t-s}^{00} \mathbf{p}_{t-s}^{10}) e^{-s} ds,$$

where the first term of the sum stands for the probability that $f(a) = 1$ and $f(b) = 0$ knowing that the Yule tree is still a single vertex at time t , and the second term is the probability of the same event, knowing that the root has given birth to two children before time t . In the second term, we look at the different possibilities to get $f(a) = 1$ and $f(b) = 0$, depending on the value of the root's label (\wedge or \vee with probability $1/2$) and on the values of the two subtrees for the assignments a and b . Simplifying, we get,

$$\mathbf{p}_t^{10} = \frac{e^{-t}}{2k} c_{a,b} + e^{-t} \int_0^t (\mathbf{p}_s^{10} - (\mathbf{p}_s^{10})^2) e^s ds$$

where $c_{a,b} = \sum_{i=1}^k \mathbb{1}_{\{a_i \neq b_i\}}$ is a constant depending only on a and b . Let $\pi_{a,b}(t) = \mathbf{p}_t^{10}(a, b)$. We have that

$$e^t \pi_{a,b}(t) = \frac{c_{a,b}}{2k} + \int_0^t (\pi_{a,b}(s) - \pi_{a,b}(s)^2) e^s ds. \quad (3)$$

Therefore we obtain the following result on $\pi_{a,b}(t)$:

Proposition 2. (i) If $a \neq b$ then $\pi_{a,b}(t) = \frac{1}{t+t_{a,b}}$ where $t_{a,b} = \frac{2k}{c_{a,b}}$.

(ii) If $a = b$, then $\pi_{a,a}(t)$ is the constant function equal to zero.

As a consequence, $\pi_{a,b}(t) = \mathbf{p}_{t,k}^{10}(a, b) \rightarrow 0$ when $t \rightarrow +\infty$ for all a, b in $\{0, 1\}^k$.

Proof. We can easily show that $\pi_{a,b}$ is differentiable. Through Equation (3) we get $\pi'_{a,b} + \pi_{a,b}^2 = 0$. If $a \neq b$, there exists $i_0 \in \{1, \dots, k\}$ such that $a_{i_0} \neq b_{i_0}$. Therefore, $c_{a,b} \geq \mathbb{1}_{\{a_{i_0} \neq b_{i_0}\}} = 1$ and $\pi_{a,b}(0) = \frac{c_{a,b}}{2k} > 0$; thus $\pi_{a,b}(t) = \frac{1}{t+t_{a,b}}$ where $t_{a,b} = \frac{2k}{c_{a,b}}$.

If $a = b$, then $\pi_{a,a}(0) = 0$ and $\pi_{a,a}(t) = 0$ for all t : a single element a cannot have two different images by a function f . \square

To obtain the convergence of $\mathbf{p}_{t,k}$ towards a limit when t tends to infinity, we only have to note that

$$\begin{aligned} \mathbf{p}_{t,k}(\mathcal{F}_k \setminus \{True, False\}) &\leq \sum_{(a,b), a \neq b} \mathbf{p}_{t,k}(f(a) = 1 \text{ and } f(b) = 0) \\ &\leq 2^k(2^k - 1) \sup_{(a,b), a \neq b} \mathbf{p}_{t,k}^{10}(a, b) \\ &\leq \frac{2^k(2^k - 1)}{t + t_{min}}, \end{aligned}$$

where $t_{min} = \inf_{(a,b), a \neq b} \{\frac{2k}{c_{a,b}}\} = 2k$. Then for all functions $f \notin \{True, False\}$, we have $\lim_{t \rightarrow +\infty} \mathbf{p}_{t,k}(f) = 0$. Moreover,

$$\mathbf{p}_{t,k}(\{True, False\}) \geq \left(1 - \frac{2^k(2^k - 1)}{t + 2k}\right),$$

which leads to $\lim_{t \rightarrow +\infty} \mathbf{p}_{t,k}(True) + \mathbf{p}_{t,k}(False) \geq 1$, i.e. $\lim_{t \rightarrow +\infty} \mathbf{p}_{t,k}(True) = \lim_{t \rightarrow +\infty} \mathbf{p}_{t,k}(False) = 1/2$. Thus, $\mathbf{p}_{t,k}$ tends to a limit distribution $p_k = \frac{1}{2}\delta_{True} + \frac{1}{2}\delta_{False}$ with a convergence speed of order $1/t$;

$$\|\mathbf{p}_{t,k} - p_k\|_\infty := \sup_{f \in \mathcal{F}_k} |\mathbf{p}_{t,k}(f) - p_k(f)| \leq \frac{2^k(2^k - 1)}{t + 2k}, \quad \text{for all } t \geq 0. \quad (4)$$

Proposition 3, which deals with the convergence speed of the sequence $(p_{n,k})$, concludes the proof of Theorem 5.

Proposition 3. *We have, asymptotically when n tends to $+\infty$, $\|p_{n,k} - p_k\|_\infty = \mathcal{O}(\frac{1}{\ln n})$.*

The proof of Proposition 3 requires the following definition and proposition.

Definition 10. *Let us define τ_n as the random variable $\tau_n = \inf\{t \geq 0 \text{ such that } N(t) = n\}$.*

Proposition 4 (see Athreya-Ney [AN72]). *For all $n \geq 1$, for all $t \geq 0$, we have,*

$$\mathbb{P}(\tau_n \leq t) = \mathbb{P}(N(t) \geq n) = (1 - e^{-t})^n.$$

Proof of Proposition 3. For all fixed $a, b \in \{0, 1\}^k$, set $\mathbb{P}_n^{10}(a, b) = \mathbb{P}(f_n(a) = 1 \text{ and } f_n(b) = 0)$, where f_n is the random Boolean function represented by $\mathcal{T}_{n,k}$, the random sprouting tree of size n . Set also, for all $t \geq 0$, η_t as the random Boolean function represented by $\mathcal{Y}_{t,k}$.

Recall that $N(t)$ is the number of internal nodes of \mathcal{Y}_t . For all $a, b \in \{0, 1\}^k$, the following equalities hold:

$$\begin{aligned} \mathbb{P}_t^{10}(a, b) &= \sum_{n \geq 0} \mathbb{P}(\eta_t(a) = 1 \text{ and } \eta_t(b) = 0 \mid N(t) = n) \mathbb{P}(N(t) = n) \\ &= \sum_{n \geq 0} \mathbb{P}(\eta_{\tau_n}(a) = 1 \text{ and } \eta_{\tau_n}(b) = 0 \mid N(t) = n) \mathbb{P}(N(t) = n). \end{aligned}$$

The embedding principle ensures that the random variables $N(t)$ and η_{τ_n} are independent for all $n \geq 0$. Since $N(t)$ follows the shifted geometric law of parameter e^{-t} , we get

$$\begin{aligned} \mathbb{P}_t^{10}(a, b) &= \sum_{n \geq 0} \mathbb{P}(\eta_{\tau_n}(a) = 1 \text{ and } \eta_{\tau_n}(b) = 0) e^{-t} (1 - e^{-t})^n \\ &= e^{-t} \sum_{n \geq 0} \mathbb{P}(f_n(a) = 1 \text{ and } f_n(b) = 0) (1 - e^{-t})^n \\ &= e^{-t} \sum_{n \geq 0} \mathbb{P}_n^{10}(a, b) (1 - e^{-t})^n. \end{aligned}$$

For all $t \geq 0$, for all $a \neq b$, in view of Proposition 2,

$$e^{-t} \sum_{n \geq 0} \mathbb{P}_n^{10}(a, b) (1 - e^{-t})^n = \frac{1}{t + t_{a,b}}.$$

Let $z = 1 - e^{-t}$; for all $z \in [0, 1[$, we denote $\varphi_{a,b}(z) = \sum_{n \geq 0} \mathbb{P}_n^{10}(a, b) z^n$. Thus,

$$\varphi_{a,b}(z) = \frac{1}{(1-z) \left(t_{a,b} + \ln \frac{1}{1-z} \right)}. \quad (5)$$

Applying a Tauberian theorem (see [FS09, page 435] for a simple statement) to this equality, we see that, when n tends to infinity, the partial sum of the first n^{th} coefficients of $\varphi_{a,b}$ is of order $\frac{n}{\ln n}$: for all $a \neq b \in \{0, 1\}^k$, asymptotically when n tends to infinity,

$$\sum_{m=0}^n \mathbb{P}_m^{10}(a, b) \sim \frac{n}{\ln n}.$$

Therefore, for large enough n , we have

$$\sum_{m=0}^n \mathbb{P}_m^{10}(a, b) \leq 2 \frac{n}{\ln n},$$

and

$$\sum_{m=0}^n p_{m,k}(f \notin \{V, F\}) \leq \sum_{a \neq b \in \{0,1\}^k} \sum_{m=0}^n \mathbb{P}_m^{10}(a, b) \leq 2^k (2^k - 1) \cdot 2 \frac{n}{\ln n}.$$

This implies that

$$\sum_{m=0}^n p_{m,k}(True) = \sum_{m=0}^n \frac{1}{2} (1 - p_{m,k}(f \notin \{T, F\})) \geq \frac{n}{2} - 2^k (2^k - 1) \frac{n}{\ln n}.$$

Finally, in view of the recurrence formula (1) applied to $f = True$,

$$\begin{aligned} p_{n+1,k}(True) &= \frac{1}{2} \sum_{g \wedge h = True} \sum_{i=0}^n \frac{1}{n+1} p_{i,k}(g) p_{n-i,k}(h) + \frac{1}{2} \sum_{g \vee h = True} \sum_{i=0}^n \frac{1}{n+1} p_{i,k}(g) p_{n-i,k}(h) \\ &\geq \frac{1}{2(n+1)} \sum_{i=0}^n p_{i,k}(True) p_{n-i,k}(True) \\ &\quad + \frac{1}{n+1} \sum_{i=0}^n p_{i,k}(True) (1 - p_{n-i,k}(True)) + \frac{1}{2(n+1)} \sum_{i=0}^n p_{i,k}(True) p_{n-i,k}(True) \\ &= \frac{1}{n+1} \sum_{i=0}^n p_{i,k}(True) \\ &\geq \frac{n}{n+1} \left(\frac{1}{2} - 2^k (2^k - 1) \frac{1}{\ln n} \right). \end{aligned}$$

and

$$\frac{1}{2} - p_{n+1,k}(True) \leq \frac{1}{2} - \frac{n}{n+1} \left(\frac{1}{2} - 2^k (2^k - 1) \frac{1}{\ln n} \right) = \mathcal{O} \left(\frac{1}{\ln n} \right),$$

which ends the proof of Proposition 3. □

4 Extensions of the And/Or sprouting tree model

In this section and in the following one, we extend our results to more general models. We bias first the distribution over the literals in both labelling models (And/Or trees and implication trees), then the law over the connectives in the And/Or model; finally, we study the And/Or model with only positive literals. These last two labelling models have been studied by Fournier *et al.* [FGG09] in the case of fully balanced binary trees, i.e. binary trees whose leaves are all at the same level, where they obtained results very similar to ours.

4.1 Biasing the distribution over literals

As before, we label each node independently by \wedge or \vee with probability $1/2$. But we now label each leaf according to a probability distribution ν , such that $\forall i \in \{1, \dots, k\}$, $\nu(x_i) = \nu(\bar{x}_i) > 0$, independently from each other. The symmetry between the variables and their negations still holds, and the behaviour of the induced probability distribution $p_{n,k}$ over \mathcal{F}_k is the same as in the case we have studied in the preceding section (when ν is the uniform law over $\{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$).

Indeed, in both proofs of Theorem 5 developed in the uniform case, the modifications appear only in constants, $c_{a,b}$ in (3) for example, which disappear when we take the derivative of the equations. Therefore, the result is the same as for the uniform case.

4.2 Biasing the distribution over connectives

We now allow further dissymmetry, between the two connectives. We define the *biased* model with parameter $\varpi \in [0, 1]$ as follows:

- each internal node is labelled independently and according to the distribution $\varpi\delta_\wedge + (1 - \varpi)\delta_\vee$ with $\varpi \in [0, 1]$, where δ_\wedge (resp. δ_\vee) is the probability distribution on the set $\{\wedge, \vee\}$ which gives probability 1 to \wedge (resp. \vee);
- each leaf is labelled independently and according to a probability distribution ν over $\{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$, such that $\forall i \in \{1, \dots, k\}$, $\nu(x_i) = \nu(\bar{x}_i) > 0$.

This process induces a new distribution over \mathcal{F}_k , still denoted by $p_{n,k}$ for simplicity's sake, whose behaviour is given by the following result:

Theorem 7. *In the biased model with parameter $\varpi \in [0, 1]$, $\varpi \neq \frac{1}{2}$,*

- *If $\varpi > \frac{1}{2}$, then $p_{n,k} \rightarrow \delta_{False}$ when n tends to infinity.*
- *If $\varpi < \frac{1}{2}$, then $p_{n,k} \rightarrow \delta_{True}$ when n tends to infinity.*

Moreover, the convergence speed is of order $\mathcal{O}\left(\frac{1}{n|2\varpi-1|}\right)$ in both cases.

Just as for connectives of equal probability, the asymptotic distribution gives a probability equal to zero to non-constant functions; however when the And connective (resp. the Or connective) is dominant, the sprouting tree distribution gives a probability equal to one to *False* (resp. *True*).

Remark: It is interesting to note that in the balanced case $\varpi = \frac{1}{2}$ (see Theorem 5) the convergence speed is of order $\frac{1}{\ln n}$, while it is of order $\frac{1}{n|2\varpi-1|}$ in Theorem 7.

Proof. We could develop two different proofs, similar to what we did when proving Theorem 5: an analytic one and a probabilistic one. We do not develop the analytic proof, and present below the probabilistic one.

The cases $\varpi > \frac{1}{2}$ and $\varpi < \frac{1}{2}$ are symmetric and can be treated in the same way. In the sequel, we only consider the case $\varpi > \frac{1}{2}$. Consider the labelled Yule tree $(\mathcal{Y}_{t,k})_{t \geq 0}$ which induces a distribution $\mathbf{p}_{t,k}$ over \mathcal{F}_k for all $t \geq 0$. Let $a = (a_1, \dots, a_k) \in \{0, 1\}^k$ be an assignment of the k variables. Set $\pi_a(t) := \mathbf{p}_{t,k}(f(a) = 1)$ as the probability that the image of a by a random Boolean function of law $\mathbf{p}_{t,k}$ is 1. We prove below that $\pi_a(t) := \mathbf{p}_{t,k}(f(a) = 1)$ tends to 0 when t tends to infinity.

Conditioning by the time when the root gives birth to two children, which is a random variable with exponential law of parameter 1, we get

$$\pi_a(t) = e^{-t} \sum_{i=1}^k (\nu(x_i) \mathbf{1}_{\{a_i=1\}} + \nu(\bar{x}_i) \mathbf{1}_{\{a_i=0\}}) + \int_0^t [\varpi \pi_a(t-s)^2 + (1-\varpi)(2\pi_a(t-s) - \pi_a(t-s)^2)] e^{-s} ds,$$

which gives

$$e^t \pi_a(t) = \frac{1}{2} + \int_0^t ((2\varpi - 1)\pi_a(s)^2 + 2(1-\varpi)\pi_a(s)) e^s ds.$$

Taking derivatives with respect to t and taking into account $\varpi \neq \frac{1}{2}$, we get $\pi_a + \pi'_a = (2\varpi - 1)\pi_a^2 + 2(1 - \varpi)\pi_a$, from which we deduce that $\pi'_a = (2\varpi - 1)(\pi_a^2 - \pi_a)$, and finally that $\pi_a(t) = \frac{1}{e^{(2\varpi-1)t} + 1}$ since $\pi_a(0) = \frac{1}{2}$. Thus

$$\mathbf{p}_{t,k}(\mathcal{F}_k \setminus \{False\}) \leq \sum_a \pi_a(t) \leq 2^k \left(\frac{1}{e^{(2\varpi-1)t} + 1} \right).$$

Since $\varpi > \frac{1}{2}$, we have $\lim_{t \rightarrow +\infty} \mathbf{p}_{t,k}(\mathcal{F}_k \setminus \{False\}) = 0$ and

$$\|p_{n,k} - \delta_{False}\|_\infty = \mathcal{O}\left(\frac{1}{n^{2\varpi-1}}\right),$$

through arguments similar to those given in the proof of Proposition 3. \square

4.3 Allowing positive literals only

Inspired again by a model studied in [FGG09] for balanced trees, we define the *positive* model as follows:

- each internal node is labelled (independently from the other nodes) according to the distribution $\varpi\delta_\wedge + (1 - \varpi)\delta_\vee$ with $\varpi \in [0, 1]$,
- each leaf is labelled independently from the other leaves according to a distribution ν over $\{x_1, \dots, x_k\}$.

The law over connectives is still biased, but we now forbid labelling the leaves by negative literals. This new model defines a new induced distribution, still denoted $p_{n,k}$, over \mathcal{F}_k , whose behaviour is given by the following theorem:

Theorem 8. *In the positive model with $\varpi \neq 1/2$, we have,*

- *If $\varpi > 1/2$, then $p_{n,k} \rightarrow \delta_{x_1 \wedge \dots \wedge x_k}$.*
- *If $\varpi < 1/2$, then $p_{n,k} \rightarrow \delta_{x_1 \vee \dots \vee x_k}$.*

The convergence speed in both cases is of order $\mathcal{O}\left(\frac{1}{n^{\lfloor 2\varpi-1 \rfloor}}\right)$.

In the positive model the sprouting tree distribution gives a nonzero probability to some non-constant function. Actually this property should be expected, as the positive model is not complete ; in particular, the constant functions *True* and *False* cannot be represented in this system.

Proof. As in the biased model, the proofs for $\varpi > 1/2$ and $\varpi < 1/2$ are very similar. We assume $\varpi > 1/2$ in the sequel. Here again, we only develop the probabilistic approach. If $a = (1, \dots, 1)$ then $\pi_a(0) = \sum_{i=1}^k \nu(x_i) \mathbb{1}_{\{a_i=1\}} = 1$ and $\pi_a(t) = 1$. Thus $\mathbf{p}_{t,k}(f(1, \dots, 1) = 1) = 1$. If $a = (0, \dots, 0)$ then $\pi_a(t) = 0$ for all $t \geq 0$. Otherwise, if $a \neq (1, \dots, 1)$ and $a \neq (0, \dots, 0)$, by the same computation as in the proof of Theorem 7, there exists a constant $c_a > 0$ such that,

$$\pi_a(t) = \mathbf{p}_{t,k}(f(a) = 1) = \frac{1}{c_a e^{(2\varpi-1)t} + 1} \text{ for all } t \geq 0,$$

and since $\varpi > \frac{1}{2}$, we have $\lim_{t \rightarrow +\infty} \mathbf{p}_{t,k}(f(a) = 1) = 0$. Thus, the asymptotic distribution of the sequence $(p_{n,k})_{n \geq 0}$ exists and gives a probability equal to one to the function $x_1 \wedge \dots \wedge x_k$. \square

Theorem 8 does not cover the case $\varpi = 1/2$, which is a natural variation of the And/Or model. Surprisingly, this last case is the most intricate of the whole study. To state our last theorem of the section, we have to present the definition of a *threshold* function: the definitions below are already stated in [Ser04] or [FGG09]. We then show that the asymptotic distribution of the $p_{n,k}$ exists and that its support is included in a finite set of threshold functions.

Definition 11. *Let $a = (a_1, \dots, a_k) \in \{0, 1\}^k$. The weight of a , relative to a distribution ν on the Boolean variables $\{x_1, \dots, x_k\}$, is the real number $\omega_\nu(a) = \nu(x_1)a_1 + \dots + \nu(x_k)a_k$.*

Definition 12. A Boolean function f is a threshold function if there exists a real number $\theta \geq 0$ such that $\forall (a_1, \dots, a_k) \in \{0, 1\}^k$, $f(a_1, \dots, a_k) = 1 \Leftrightarrow \omega_\nu(a) \geq \theta$. We denote by $f_{\nu, \theta}$ the threshold function associated with the constant θ and the distribution ν .

Theorem 9. Order the elements of $\{0, 1\}^k$ by increasing weight ω_ν (see Definition 11): $\omega_\nu(a^{(1)}) \leq \omega_\nu(a^{(2)}) \leq \dots \leq \omega_\nu(a^{(2^k)})$. Then,

$$p_{n,k} \longrightarrow \sum_{j=2}^{2^k} \left(\omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)}) \right) \delta_{f_{\nu, \omega_\nu(a^{(j)})}} \text{ when } n \rightarrow \infty.$$

In other words, $p_{n,k}$ tends to the asymptotic distribution p_k that satisfies

$$p_k(f_{\nu, \omega_\nu(a^{(j)})}) = \omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)}) \text{ for all } j \in \{2, \dots, 2^k\},$$

and if f is a Boolean function different from $f_{\nu, \omega_\nu(a^{(j)})}$ for all $j \in \{2, \dots, 2^k\}$, then $p_k(f) = 0$.

Proof. The proof is once again based on Yule trees. The probabilistic approach is the most natural in this case, since it is an extension of the proof developed in [FGG09] in the case of balanced trees. Let $\mathcal{Y}_{t,k}$ be the labelled Yule tree, $a = (a_1, \dots, a_k)$ and $b = (b_1, \dots, b_k)$ in $\{0, 1\}^k$ two assignments of the k variables, α and β in $\{0, 1\}$. For all $t \geq 0$ set

$$\pi_{\alpha\beta}(t) = \mathbf{p}_{t,k}(f(a) = \alpha \text{ and } f(b) = \beta).$$

We compute π_{10} by conditioning on the time when the root gives birth to two children.

$$\begin{aligned} \pi_{10}(t) = \mathbf{e}^{-t} \sum_{i=1}^k a_i(1 - b_i) \nu(x_i) + \int_0^t \frac{1}{2} [\pi_{11}(t-s)\pi_{10}(t-s) + \pi_{10}(t-s)(\pi_{10}(t-s) + \pi_{11}(t-s)) \\ + \pi_{10}(t-s)(\pi_{10}(t-s) + \pi_{00}(t-s)) + \pi_{10}(t-s)\pi_{00}(t-s)] \mathbf{e}^{-s} ds \end{aligned}$$

This gives

$$\pi_{10}(t) \mathbf{e}^t = \sum_{i=1}^k a_i(1 - b_i) \nu(x_i) + \int_0^t (\pi_{10}(s)^2 + \pi_{10}(s)\pi_{11}(s) + \pi_{10}(s)\pi_{00}(s)) \mathbf{e}^s ds.$$

By differentiating and using the (obvious) relation $\pi_{11} + \pi_{10} + \pi_{01} + \pi_{00} = 1$, we get $\pi'_{10} = -\pi_{10}\pi_{01}$. Doing the same computation for π_{00} , π_{01} and π_{11} leads to the differential system:

$$\begin{cases} \pi'_{10} &= -\pi_{10}\pi_{01}; \\ \pi'_{01} &= -\pi_{10}\pi_{01}; \\ \pi'_{11} &= \pi_{10}\pi_{01}; \\ \pi'_{00} &= \pi_{10}\pi_{01}. \end{cases} \quad (6)$$

The first two equations of (6) imply that $\pi_{10}(t)$ and $\pi_{01}(t)$ are decreasing functions of t ; since they are both positive, they have a limit as $t \rightarrow +\infty$. In the same vein we check that π_{11} and π_{00} are increasing, lower than 1, and thus convergent. Let us denote $l_{\alpha\beta} = \lim_{t \rightarrow +\infty} \pi_{\alpha\beta}(t)$.

Since $\pi_{\alpha\beta}$ is monotone and convergent for $t \rightarrow +\infty$, its derivative tends to zero as $t \rightarrow +\infty$. Thus, taking the limit in (6), we get

$$l_{10}l_{01} = 0. \quad (7)$$

Moreover we deduce from (6) that $\pi_{10} - \pi_{01}$ is a constant, which implies that

$$l_{10} - l_{01} = \pi_{10}(0) - \pi_{01}(0) = \omega_\nu(a) - \omega_\nu(b). \quad (8)$$

Thus if $\omega_\nu(a) \geq \omega_\nu(b)$, then by applying equations (7) and (8) we get $l_{01} = 0$. This means that $\mathbf{p}_{t,k}(f(a) = 0 \text{ and } f(b) = 1) \rightarrow 0$ as $t \rightarrow +\infty$. Said differently, if there exist a and b such that $\omega_\nu(a) \geq \omega_\nu(b)$ and $f(a) = 0$ and $f(b) = 1$, then $p_{n,k}(f) \rightarrow 0$ as $n \rightarrow +\infty$. Hence, the only Boolean functions weighted by $p_{n,k}$ when n

tends to infinity are those verifying the following property: for all a, b such that $\omega_\nu(a) \geq \omega_\nu(b)$, $f(a) \geq f(b)$. Such functions are threshold functions; if the asymptotic distribution exists when n tends to infinity, then its support is contained in the set of threshold functions.

The calculations we made in the non-uniform positive model can be duplicated here to prove that $\mathbf{p}_{t,k}(f(a) = 1)$ is constant for all a . Thus $\mathbf{p}_{t,k}(f(a) = 1) = \omega_\nu(a)$ and for all $j \in \{2, \dots, 2^k\}$,

$$p_{n,k}(f_{\nu, \omega_\nu(a^{(1)})}) + \dots + p_{n,k}(f_{\nu, \omega_\nu(a^{(j)})}) \rightarrow \omega_\nu(a^{(j)}).$$

Thus $p_{n,k}(f_{\nu, \omega_\nu(a^{(j)})}) \rightarrow \omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)})$ when $n \rightarrow +\infty$, and as $\sum_{j=2}^{2^k} \omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)}) = 1$, we have indeed proved Theorem 9. \square

5 Tautologies in the implication model

In this section we set aside the And/Or model and concentrate on the implication model, in which the internal nodes of the trees under consideration are labelled by implication symbols (\rightarrow). We now allow only positive literals as labels for the leaves in the first subsection, whereas positive and negative literals are allowed in the second subsection.

5.1 The positive implication model

As mentioned in the Introduction, this model has been the subject of several papers under both the Catalan tree and the Galton-Watson tree models. Most notably, tautological trees (i.e. trees which represent the identically *True* function) have been closely investigated. E.g., Fournier *et al.* ([FGGZ07]) proved that the classical and intuitionistic logics are *asymptotically identical*, or studied ([FGGG12]) the relation between the probability of a function and its complexity. It was shown there that, for both the Catalan tree and the Galton-Watson tree models, *asymptotically almost all tautologies are simple*. The aim of this section is to understand whether this property still holds for the sprouting tree distribution.

First of all, we define *simple* tautologies.

Definition 13 ([FGGZ07]). *In the implication labelling model, every Boolean expression can be written as $A_1 \rightarrow (A_2 \rightarrow \dots (A_p \rightarrow \alpha) \dots)$ where the $(A_i)_{i=1, \dots, p}$ are implication Boolean expressions and α is a literal. The subtrees representing A_1, \dots, A_p are called the premises of the Boolean tree and α is called the goal. A simple tautology is a Boolean tree which has at least one premise reduced to a simple leaf labelled by α (see Figure 2).*

We denote by $ST_{n,k}$ the set of simple tautologies of size n over k variables.

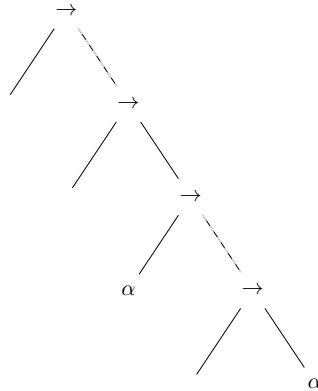


FIGURE 2 – Simple tautology

The following theorem holds for both the Catalan tree and the Galton-Watson tree distributions:

Theorem 10 (Fournier *et al.* [FGGZ07, FGGG12]). *The probabilities $\mu_k(\text{True})$ and $\pi_k(\text{True})$ of the function identically True with respect to the Catalan tree distribution and to the Galton-Watson tree distribution satisfy respectively*

- $\mu_k(\text{True}) \sim \lim_{n \rightarrow +\infty} \frac{|ST_{n,k}|}{T_{n,k}}$ asymptotically when k tends to infinity, where $T_{n,k}$ is the total number of binary trees of size n labelled with k variables; and
- $\pi_k(\text{True}) \sim \Pi_k(ST_k)$ asymptotically when k tends to infinity, where $\Pi_k(ST_k)$ is the probability that the labelled Galton-Watson tree is a simple tautology.

Does the sprouting tree distribution have the same kind of behaviour? This section allows us to answer negatively.

Theorem 11. *We have $\mathbb{P}_{n,k}(ST_{n,k}) \xrightarrow{n \rightarrow +\infty} 1 - e^{-1/k} \sim 1/k$ when $k \rightarrow +\infty$.*

As a consequence, since the probability of *True* with respect to the sprouting tree distribution is 1 (see Theorem 6), then asymptotically when k tends to infinity, almost no tautology is simple. Once again, the behaviour of the sprouting tree distribution is different from that of Catalan tree and Galton-Watson tree distributions.

The proof of the theorem uses some knowledge about Pólya urns and their study by analytic combinatorics (see [FGP05] for details on this topic).

Proof. The proof comprises two steps: we calculate first the distribution of the number f_n of premises that are reduced to a simple leaf in a sprouting tree of size n (we call them *nice premises*), then the probability to get a simple tautology, by conditioning over the number of nice premises.

The first step can be handled by modelling the system with a Pólya urn. Indeed, let us consider an urn containing three kinds of balls, representing three possible kinds of leaves of the tree: the white balls stand for the nice premises, a single red ball stands for the goal of the Boolean tree, and the black balls stand for the other leaves. When the tree grows, we choose one of its leaves (i.e. one of the balls) uniformly at random, and

- if we choose the red ball, then we put it back into the urn and add a white ball (i.e. a nice premise);
- if we choose a white ball, then we remove it from the urn and add two black balls into the urn;
- if we choose a black ball, then we put it back into the urn and add another black ball.

The replacement matrix of this urn is given by

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

and its initial composition is a single red ball. At time n , the urn contains $n + 1$ balls and represents the sprouting tree of size n . Its behaviour has been studied by Morcrette [Mor10] who proved that the number of white balls in the urn at time n , i.e. the number f_n of nice premises in the sprouting tree at time n follows the same distribution as the number of fixed points in a uniformly random permutation of length n and satisfies, for all $n \geq 1$ and all $m \leq n$, the equality

$$\mathbb{P}_{n,k}(f_n = m) = \frac{1}{m!} \left(e^{-1} - \sum_{j \geq n+1-m} \frac{(-1)^j}{j!} \right) = \frac{1}{m!} \sum_{j=0}^{n-m} \frac{(-1)^j}{j!}. \quad (9)$$

Let us now evaluate $\mathbb{P}_{n,k}(ST_{n,k})$ by conditioning on the number m of nice premises in the sprouting tree. Since $(1 - (1 - \frac{1}{k})^m)$ is the probability that one of the m nice premises is labelled by the same label as the goal of the Boolean tree, we get

$$\mathbb{P}_{n,k}(ST_{n,k}) = \sum_{m=1}^n \mathbb{P}_{n,k}(f_n = m) \left(1 - \left(1 - \frac{1}{k} \right)^m \right).$$

Setting $c = 1 - \frac{1}{k}$ and plugging the value of $\mathbb{P}_{n,k}(ST_{n,k})$ that we have just obtained, we get

$$\mathbb{P}_{n,k}(ST_{n,k}) = \sum_{m=1}^n \frac{(1-c^m)}{m!} \sum_{j=0}^{n-m} \frac{(-1)^j}{j!} = \sum_{s=1}^n \sum_{j+m=s} \frac{(1-c^m)}{m!} \frac{(-1)^j}{j!}.$$

This series converges to

$$e^{-1}(e^1 - e^c) = 1 - e^{c-1},$$

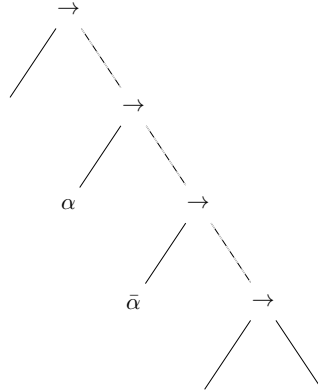
which concludes the proof since $c = 1 - \frac{1}{k}$. \square

5.2 Allowing positive and negative literals

Simple tautologies have also been studied under a variation of the implication model, where negative literals are allowed [FGGZ10]. As for the classical implication model, we can prove that the sprouting tree distribution $p_{n,k}$ tends to δ_{True} when n tends to infinity. In this new labelling model, there are two kinds of simple tautologies: simple tautologies of first kind, defined in the same way as in the classic labelling model (see Definition 13), and simple tautologies of second kind, which we now define:

Definition 14 ([FGGZ10]). *A tautology of second kind is a Boolean expression in which two nice premises are labelled respectively with a variable and its negation (see Figure 3). We denote by $ST_{n,k}^1$ (resp. $ST_{n,k}^2$) the set of simple tautologies of first kind (resp. second kind).*

FIGURE 3 – Simple tautology of the second kind



Fournier *et al.* ([FGGZ10]) have shown that, both in the Catalan and in the Galton-Watson models with positive and negative literals, all the tautologies are simple tautologies of either first or second kind, asymptotically when k tends to infinity. We prove that this is not the case in the sprouting tree model.

Theorem 12. *We have*

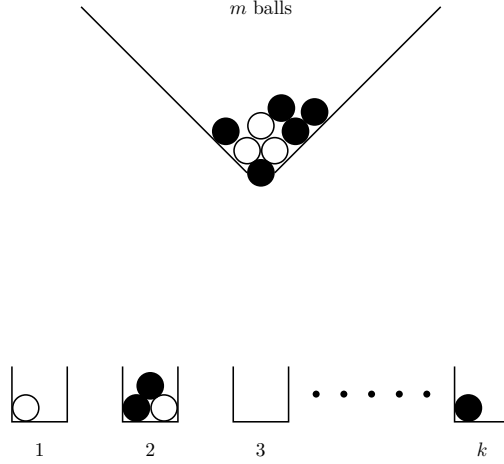
- $\mathbb{P}_{n,k}(ST_{n,k}^1) \xrightarrow{n \rightarrow +\infty} 1 - e^{-1/2k} \sim \frac{1}{2k}$, when $k \rightarrow +\infty$; and
- $\mathbb{P}_{n,k}(ST_{n,k}^2) \xrightarrow{n \rightarrow +\infty} 1 - \frac{1}{e}(2e^{1/2k} - 1)^k \sim \frac{1}{4k}$, when $k \rightarrow +\infty$.

Since the probability of *True* is 1 under the sprouting tree distribution in the implication model (see Theorem 6), it follows that, asymptotically when k tends to infinity, almost no tautology is simple.

The proof of this theorem relies on the study of some allocation problem. Such topics are treated for example in [JK97]; we use here an analytic combinatorics approach (we refer to [Gar02] for a survey about this method).

Proof. The first statement of Theorem 12 can be shown in the same way as Theorem 11. Let us consider the second one. Equation (9) still holds and we only have to compute the probability that two nice premises among m are labelled by a variable and its negation. We express this problem as a birthday problem [JK97] as follows. We assign m balls (labels of the nice premises) into k urns (Boolean variables). We color each ball independently, either white or black (positive or negative literals) with probability $1/2$. The probability that at least one urn contains at least one black ball and one white ball is the probability of a simple tautology of second kind (see Figure 4).

FIGURE 4 – Allocation problem associated with the study of simple tautologies of the second kind



This allocation problem is solved through the use of generating functions as follows. Consider a single urn: the generating function e^{x+y} counts the number of different allocations of balls into the urn, with x marking the positive occurrences of the Boolean variable associated with the urn and y its negative occurrences. We can decompose this generating function as

$$e^{x+y} = (e^x - 1)(e^y - 1) + (e^x - 1) + (e^y - 1) + 1,$$

where the first term counts the allocations where both the variable and its negation appear, the second one the allocations where only the variable appears, the third one those where only the negation of the variable appears, and the last one stands for an empty urn. Introduce a new variable z that marks the joint occurrence of both the Boolean variable and its negation; the generating function of the different allocations becomes

$$z(e^x - 1)(e^y - 1) + (e^x - 1) + (e^y - 1) + 1.$$

We now “forget” the difference between a variable and its negation; x and y both become t and the description of the possible states for a single urn is given by the following generating function

$$z(e^t - 1)^2 + 2e^t - 1.$$

Therefore, if $\alpha_{r,m}$ is the number of assignments of m balls into k urns that leads in r different ways to a simple tautology of second kind, then

$$\Phi(t, z) := \sum_{r,m} \alpha_{r,m} z^r \frac{t^m}{m!} = (z(e^t - 1)^2 + 2e^t - 1)^k.$$

As $\Phi(t, 0)$ is the generating function of the number of assignments of the m balls that do *not* lead to a simple tautology of second kind, we obtain

$$\mathbb{P}_{n,k}(\overline{ST_{n,k}^2} | f_n = m) = \frac{\frac{t^m}{m!} \Phi(t, 0)}{(2k)^m} = \frac{\alpha_{0,m}}{(2k)^m},$$

which with (9) gives

$$\mathbb{P}_{n,k}(\overline{ST_{n,k}^2}) = \underbrace{e^{-1} \sum_{m=0}^n \frac{\alpha_{0,m}}{m! (2k)^m}}_{Q_n} - \underbrace{\sum_{m=0}^n \frac{\alpha_{0,m}}{m! (2k)^m} \sum_{j \geq n+1-m} \frac{(-1)^j}{j!}}_{R_n}.$$

We can easily prove that R_n tends to zero as n tends to infinity, and that

$$Q_n \rightarrow \frac{1}{e} \Phi\left(\frac{1}{2k}, 0\right) \text{ when } n \rightarrow +\infty.$$

Thus

$$\mathbb{P}_{n,k}(\overline{ST_{n,k}^2}) \xrightarrow{n \rightarrow +\infty} \frac{1}{e} \Phi\left(\frac{1}{2k}, 0\right) = \frac{1}{e} (2e^{1/2k} - 1)^k \sim 1 - \frac{1}{4k}$$

when $k \rightarrow +\infty$. □

6 Conclusions and perspectives

We were able to obtain a complete overview of the behaviour of the sprouting tree distribution over the set of Boolean functions. The methods we used were quite varied: from analytic combinatorics to embedding in continuous time, via Pólya urns and random allocation problems.

We have shown that this new distribution over \mathcal{F}_k has an unexpected behaviour, and differs widely from the Catalan and Galton-Watson distributions (see Table 2 for a comparison of the three models). This distribution, defined by a natural growing process inspired from the random Binary Search Trees, is degenerate; it gives probability zero to non-constant functions, both for And/Or trees and for implication trees (see Theorems 5 and 6).

	Catalan tree	Galton-Watson tree	Sprouting tree
Distribution on \mathcal{E}_k	$\mathbb{U}_{n,k}$	Π_k	$\mathbb{P}_{n,k}$
Induced distribution on \mathcal{F}_k	$\mu_{n,k}$ $\downarrow (n \rightarrow \infty)$ μ_k	π_k	$p_{n,k}$ $\downarrow (n \rightarrow \infty)$ p_k
General behaviour	$\mu_k(f) = \Theta\left(\frac{1}{k^{L(f)+2}}\right)$	$\pi_k(f) = \Theta\left(\frac{1}{k^{L(f)+1}}\right)$	$p_k(f) = \frac{1}{2}\delta_{True} + \frac{1}{2}\delta_{False}$ (And/Or) $p_k(f) = \delta_{True}$ (Implication)

TABLE 2 – *Summary of the different notations and comparison between the different models mentioned in the present paper*

In the special case of implication trees, we completed the study of the model by considering simple tautologies, which is of interest in quantitative logic; the sprouting tree distribution again presents an unexpected behaviour, since only a negligible part of tautological trees are simple tautologies when the number of variables tends to infinity (see Theorems 11 and 12), as opposed to the Catalan tree and Galton-Watson tree distributions.

A natural question then arises: why does the sprouting tree distribution have a degenerate behaviour, widely different from the behaviour of the Catalan or Galton-Watson tree distributions (and from the uniform distribution)? Which property of the random trees can explain these differences?

An empirical explanation can be guessed from Section 4: we saw there that the sprouting tree distribution behaves as the fully balanced trees distribution studied in [FGG09] (recall that a balanced tree is a complete binary tree whose leaves are all at the same level). Indeed, Theorem 7, 8 and 9 still hold for fully balanced trees; even more, due to the continuous time embedding of the sprouting tree, the proofs developed in the present paper and the proofs of [FGG09] are very similar. In a forthcoming paper, Broutin and Mailler

enlighten this similarity by proving that any *saturated* tree (tree whose minimal distance between a leaf and the root tends to infinity in probability when the size of the tree tends to infinity) induces a degenerate asymptotic distribution on Boolean functions.

Acknowledgements

We are grateful to Michael Drmota for fruitful discussions about this work and to the anonymous referees for their help in improving the present paper.

References

- [AN72] K. B. Athreya and P. E. Ney. *Branching Processes*. Springer-Verlag, 1972.
- [CFGG04] B. Chauvin, Ph. Flajolet, D. Gardy, and B. Gittenberger. And/or trees revisited. *Combinatorics, Probability and Computing*, 13(4-5):475–497, July-September 2004.
- [CKMR05] B. Chauvin, T. Klein, J.-F. Marckert, and A. Rouault. Martingales and profile of binary search trees. *Electronic Journal of Probability*, 10(12):420–435, 2005.
- [CLR89] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. The MIT Press and McGraw-Hill Book Company, 1989.
- [Drm97] M. Drmota. Systems of functional equations. *Random Structures and Algorithms*, 10(1-2):103–124, 1997.
- [FGG09] H. Fournier, D. Gardy, and A. Genitrini. Balanced and/or trees and linear threshold functions. In *5th SIAM Workshop on Analytic and Combinatorics (ANALCO)*, pages 51–57, 2009.
- [FGGG08] H. Fournier, D. Gardy, A. Genitrini, and B. Gittenberger. Complexity and limiting ratio of boolean functions over implication. In *33rd International Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 347–362, Torun, Poland, August 2008. LNCS 5162.
- [FGGG12] H. Fournier, D. Gardy, A. Genitrini, and B. Gittenberger. The fraction of large random trees representing a given Boolean function in implicational logic. *Random Structures and Algorithms*, 20(7):875–887, 2012.
- [FGGZ07] H. Fournier, D. Gardy, A. Genitrini, and M. Zaionc. Classical and intuitionistic logics are asymptotically identical. In *Proc. 16th Annual Conference on Computer Science Logic (EACSL)*, pages 177–193. LNCS 4646, 2007.
- [FGGZ10] H. Fournier, D. Gardy, A. Genitrini, and M. Zaionc. Simple tautologies over implication with negative literal. *Journal Mathematical Logic Quarterly*, 56(4):388–396, 2010.
- [FGP05] Ph. Flajolet, J. Gabarró, and H. Pekari. Analytic urns. *Annals of Probability*, 33(3):1200–1233, 2005.
- [FO90] Ph. Flajolet and A. M. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3(2):216–240, 1990.
- [FS09] Ph. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.
- [Gar02] D. Gardy. Occupancy urn models in analysis of algorithms. *Journal of Statistical Planning and Inference, special issue on the Fourth International Conference on Lattice Paths Combinatorics and Applications*, 101(1-2):95–105, February 2002.
- [Gar06] D. Gardy. Random boolean expressions. In *Proc. Colloquium on Computational Logic and Applications*, volume AF, pages 1–36. DMTCS Proceedings, 2006.

- [GG10] A. Genitrini and B. Gittenberger. No Shannon effect on probability distributions on Boolean functions induced by Boolean expressions. In *Proc. of Analysis of Algorithms*, volume AM, pages 303–316, Wien, Austria, July 2010. DMTCS Proceedings.
- [GK09] A. Genitrini and J. Kozik. Quantitative comparison of intuitionistic and classical logics - full propositional system. In *Proc. of International Symposium on Logical Foundations of Computer Science (LFCS'09)*, pages 280–294, Florida, USA, January 2009. LNCS 5407.
- [Grö19] T. H. Grönwall. Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Annals of Mathematics*, pages 292–296, 1919.
- [Har49] G. H. Hardy. *Divergent Series*. Oxford University Press, 1949.
- [JK97] N. L. Johnson and S. Kotz. *Urn Models and their Application*. Wiley and Sons, 1997.
- [Juk12] S. Jukna. *Boolean Function Complexity: Advances and Frontiers*. Springer, 2012.
- [Koz08] J. Kozik. Subcritical pattern languages for And/Or trees. In *Fifth Colloquium on Mathematics and Computer Science*, Blaubeuren, Germany, september 2008. DMTCS Proceedings.
- [KZ04] Z. Kostrzycka and M. Zaionc. Statistics of intuitionistic versus classical logic. *Studia Logica*, 76(3):307–328, 2004.
- [Lal93] S. P. Lalley. Finite range random walk on free groups and homogeneous trees. *The Annals of Probability*, 21(4):2087–2130, 1993.
- [LS97] H. Lefmann and P. Savický. Some typical properties of large and/or Boolean formulas. *Random Structures and Algorithms*, 10:337–351, 1997.
- [Lup60] O. B. Lupanov. On the complexity of realization of functions of propositional calculus by formulas. *Problemy Kibernetiki*, 3:61–80, 1960. *English translation: Problems of Cybernetics, Pergamon Press*, 3:61–80, 1960.
- [Mor10] B. Morcrette. Combinatoire analytique et modèles d’urnes. Master’s thesis, MPRI - Inria Rocquencourt, 2010. http://algo.inria.fr/morcrette/rapport_M2.pdf.
- [MTZ00] M. Moczurad, J. Tyszkiewicz, and M. Zaionc. Statistical properties of simple types. *Mathematical Structures in Computer Science*, 10(5):575–594, 2000.
- [Pit84] B. Pittel. On growing random binary trees. *Journal of Mathematical Analysis and Applications*, 103(2):461–480, 1984.
- [PVW94] J. B. Paris, A. Vencovská, and G. M. Wilmers. A natural prior probability distribution derived from the propositional calculus. *Annals of Pure and Applied Logic*, 70:243–285, 1994.
- [RS42] J. Riordan and C. E. Shannon. The number of two terminal series-parallel networks. *Journal of Math. Physics*, 21:83–93, 1942.
- [Ser04] R. A. Servedio. Monotone Boolean formulas can approximate monotone linear threshold functions. *Discrete Applied Mathematics*, 142(1-3):181–187, 2004.
- [Weg05] I. Wegener. *Complexity Theory: Exploring the Limits of Efficient Algorithms*. Springer Verlag, 2005.
- [Woo97] A. R. Woods. Coloring rules for finite trees, and probabilities of monadic second order sentences. *Random Structures and Algorithms*, 10(4):453–485, 1997.